# Линейная регрессия на плоскости.

Рассмотрим модель линейной регрессии, при которой целевая переменная зависит лишь от одного признака, тогда функция, описывающая зависимость $y$ от $x$ будет иметь следующий вид:

$$f(x) = w_0 + w_1 * x$$

и задача сводится к нахождению весовых коэффициентов $w_0$ и $w_1$, таких что такая прямая максимально "хорошо" будет описывать исходные данные. Для этого будем использовать уже обсуждённую функцию ошибки, минимизация которой обеспечит подбор весов $w_0$ и $w_1$:

$$MSE = \frac{1}{n} * \sum_{i=0}^{n} (y_i - f(x_i))^2$$

или подставив уравнение модели

$$MSE = \frac{1}{n} * \sum_{i=0}^{n} (y_i - w_0 - w_1 * x_i)^2$$

Минимизируем функцию ошибки MSE найдя частные производные по $w_0$ и $w_1$

$$\frac{\partial MSE(w_0, w_1)}{\partial w_0} = -\frac{2}{n} * \sum_{i=0}^{n} (y_i - w_0 - w_1 * x_i)$$

$$\frac{\partial MSE(w_0, w_1)}{\partial w_1} = -\frac{2}{n} * \sum_{i=0}^{n} ((y_i - w_0 - w_1 * x_i) * x_i)$$

И приравняв их к нулю получим систему уравнений, решение которой обеспечит минимизацию функции потерь MSE.

$$\begin{cases} 0 = -\frac{2}{n} * \sum\limits_{i=0}^{n} (y_i - w_0 - w_1 * x_i) \\ 0 = -\frac{2}{n} * \sum\limits_{i=0}^{n} ((y_i - w_0 - w_1 * x_i) * x_i) \end{cases}$$

Раскроем сумму и с учетом того, что $-2/n$ не может равняться нулю, приравняем к нулю вторые множители

$$\begin{cases} 0 = -w_0 * n + \sum\limits_{i=0}^{n} y_i - w_1 * \sum\limits_{i=0}^{n} x_i \\ 0 = \sum\limits_{i=0}^{n} (y_i * x_i) - w_0 * \sum\limits_{i=0}^{n} x_i - w_1 * \sum\limits_{i=0}^{n} x_i^2 \end{cases}$$

Выразим $w_0$ из первого уравнения

$$w_0 = \frac{\sum_{i=0}^{n} y_i}{n} - w_1 \frac{\sum_{i=0}^{n} x_i}{n}$$

Подставив во второе уравнение решим относительно $w_1$

$$0 = \sum_{i=0}^{n}(y_i * x_i) - (\frac{\sum_{i=0}^{n} y_i}{n} - w_1 \frac{\sum_{i=0}^{n} x_i}{n}) * \sum_{i=0}^{n} x_i - w_1 * \sum_{i=0}^{n} x_i^2$$

$$0 = \sum_{i=0}^{n}(y_i * x_i) - \frac{\sum_{i=0}^{n}(y_i \sum_{i=0}^{n} x_i)}{n} + w_1 \frac{\sum_{i=0}^{n}(x_i \sum_{i=0}^{n} x_i)}{n} - w_1 * \sum_{i=0}^{n} x_i^2$$

И выразив $w_1$ последнего уравнения получим

$$w_1 = \frac{\frac{\sum_{i=0}^{n}(x_i \sum_{i=0}^{n} y_i)}{n} - \sum_{i=0}^{n}(y_i * x_i)}{\frac{\sum_{i=0}^{n}(x_i \sum_{i=0}^{n} x_i)}{n} - \sum_{i=0}^{n} x_i^2}$$

Задача решена, однако представленный способ слабо распространим на большое количество фичей, уже при появлении второго признака вывод становится достаточно громоздким, не говоря уже о большем количестве признаков.

Основы линейной регрессии / Хабр